



Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at <http://about.jstor.org/participate-jstor/individuals/early-journal-content>.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact support@jstor.org.

A STUDY OF THE ACCURACY OF THE PRESENT METHODS OF TESTING FATIGUE.

By Professor A. CASWELL ELLIS, University of Texas, and MAUD
MARGARET SHIPE.

A variety of methods of testing fatigue have been applied to school children in recent years, and volumes of results reported ; but there has been nowhere a proper testing of the validity of these methods as tests of fatigue. Leuba, and others, have shown that the æsthesiometer test is worthless.¹ Bolton has reported experiments which indicate that the ergograph method is not accurate.² Thorndike, in a series of tests, both mental and physical (including addition, multiplication, tests of attention, memory, etc., and the dynamometer), has failed to find fatigue indicated by these tests at times when it seemed that fatigue must certainly be present.³ But a full comparative test of all the methods, with a view of testing the validity of the methods themselves, has not been reported ; and the ergograph, the reaction time, the adding, the memorizing of nonsense syllables, the filling of blanks in printed matter, are all still used as means of testing fatigue, and the results supposed to be of value. With a view of testing these methods of testing fatigue, the following experiments were undertaken.

I.

A group of five serious, advanced students and one professor was tested twice a day for five days, using more than one method of testing at each period. The tests were all short and the labor of them was not arduous enough in itself to increase to any considerable extent the fatigue present at the beginning of the test, so that practically the same condition of

¹*Psychological Review*, VI., 573 and 599.

²*Psychological Review*, VII, 136.

³*Psychological Review*, VII, 466 and 547.

fatigue was present through all the tests made at one period. The time required for the use of four tests at one period was about twenty minutes.

If these tests are accurate measures of the fatigue present, there should be agreement in results among the several methods employed at the same time. If the different tests should fail to agree in results: for instance, if in the morning test different conditions of fatigue were indicated by the different tests used at that period, or if in the afternoon tests, one test should show increased power over the morning test, another decreased power, and a third unaltered power, then some, or possibly all, of the methods must be inaccurate.

The first tests compared were the reaction time and ergograph tests. For the reaction time, the ordinary drop shutter and Hipp chronoscope were used. The reaction time tested was for the simple recognition of familiar names, such as Gray, Hume, Pope, printed in large letters on white cards. Each name contained four letters, hence the letters occupied the same space on the card in each case.

The ergograph was after the pattern of Mosso's, with an endless tape to record the total movement of the weight, and a lever writing on a smoked kymograph drum to record the number and variety of contractions. The subjects all made uniformly rapid contractions, keeping time to the beat of a metronome. Each subject was required to crook his finger till it touched a wooden stop which was placed so that it would be touched as the finger was almost completely crooked. The amplitude of the contractions thus varied with the size of the subjects' hands, but was fairly constant for each subject. The subject was required to continue his contractions till he could no longer bring his finger up to this stop. The weight was, of course, kept constant for each subject during the entire series of experiments.

The experiments were painful, but all the subjects entered into them earnestly — certainly, far more faithfully than would ordinary students when being tested for fatigue.

After a short preliminary practice to familiarize the subjects with the apparatus and the tests, a series of five tests on alternate days was made; the first test between eight and nine in

the morning, before any class work was done, and the second between twelve and one o'clock after a morning of work.

At each test the reaction time was taken first. Seven reactions, exclusive of "false alarms," were made at each sitting. The last five only were counted each time in computing the average time and the mean variation. The outside conditions were kept as favorable and as uniform as possible.

It was supposed that increased reaction time and increased variation would go together and would accompany increased fatigue and decreased power as shown by the ergograph. The experiments resulted otherwise. Of 27 perfect tests only 13 times did average reaction time and mean variation of reaction time increase or decrease together; only 16 times did the mean variation in reaction time and power, as shown by the ergograph, agree in results (*i. e.*, wide variation accompany small power, or *vice versa*); only 10 times did shortened reaction time accompany increased power as shown by the ergograph, or *vice versa*. This comparison seems to show that one or both of these tests must be inaccurate.¹

Taking separately the results from each form of test, it was found that out of 27 perfect tests, 14 times was the mean variation for reaction time less, 12 times the reaction time shorter, 13 times the power as shown by the ergograph greater at the end than at the beginning of the morning. Only 6 times did less variation, shorter reaction time, and greater power with the ergograph, or their opposite, all occur at the same test together. The tests seemed to show that the students were as fresh at noon as at the beginning of the day. The appearance and feeling of the students belied the results.

In order to try the tests with yet more severe fatigue, the same group was tested one day at 8 A. M., and again at 5 P. M., after unremitting labor except during the dinner hour. Of these 6 tests, 3 times the variation in reaction time was less, 3 times the ergograph showed more power, and 3 times the re-

¹Complete tabular statements of the detailed results in this and other cases accompanied the manuscript of this article as originally submitted by the authors. The negative thesis which the article supports seems, however, so amply demonstrated by the summaries in the text that, on the recommendation of the editors, the full tables have not been inserted.

action time was shorter in the afternoon than in the morning ; 3 times the three tests agreed in results.

II.

These methods seemed so unreliable as tests of fatigue that it was decided to add others with less motor and more intellectual elements. In the new series the ergograph was dropped, as it seemed no test of fatigue and was troublesome to use. The same subjects were used as before. Four forms of test were used : (1) reaction time (taken as before), (2) addition of columns of figures, (3) writing the cubes of numbers up to 9, (4) memorizing nonsense syllables.

The figures were printed in columns of three figures, each 27 figures deep, thus making it a task of adding units, tens, and hundreds. Both the total number of figures added and the number of columns added correctly were counted. In order that the addition of each column might be considered separately, as well as the sum as a whole, the sum of each column was written beneath it, and not just the right hand figure as is customary. In this way the amount "carried" each time was indicated, and a mistake made in one column, which caused an incorrect result in the next column, even though that was added correctly, could be discovered.

The figures to be cubed were printed in vertical lines, and the subject wrote the cube in a blank space at the right of each figure, all multiplication being mental. After a few tests it was a mere matter of memory and manual dexterity. Both the number of cubes written at a test and the accuracy of results were considered.

The nonsense syllables, containing three letters each, were printed in horizontal lines, ten in a line. The subject was required to memorize these in the order printed, and at the end of the period to turn the paper over and write, in correct order, the syllables remembered, and to state what number he had attempted to memorize. In tabulating these results the following points were considered : (1) the number attempted, (2) the number learned correctly, (3) the number incorrectly recalled, (4) the number omitted, (5) the number transposed.

In taking the last three tests, all were seated around a large

table with everything in readiness. All began and stopped at a given signal. Two minutes were allowed for adding, when papers were changed and two minutes given for cubing. After a change of papers two minutes were given for learning nonsense syllables, and finally, two minutes allowed for writing down the syllables remembered. Tests were taken at 8.30 A. M., 12.30 and 5.30 P. M., each day for four days. The time chosen was examination week, when all students were working very hard and were under fatiguing mental worry.

The results of these experiments were equally disappointing. Comparing the 8.30 A. M. and 5.30 P. M. tests, we found the following results : Of 24 perfect records for each test, the reaction time and variation increased or decreased together 13 times ; the reaction time test agreed with the addition test 13 times (*i. e.*, short reaction time accompanied a large number of figures added, or *vice versa*), with the cubing test 10 times, with the memorizing-nonsense-syllables test 6 times ; the mean variation in reaction time agreed with the adding test 11 times, with the cubing test 10 times, and with the nonsense test 8 times ; the adding and the nonsense tests agreed 8 times ; the adding and cubing tests agreed 10 times ; the cubing and the nonsense tests agreed 11 times.

Comparing now not the total work done, but the total done correctly at 8.30 A. M. and 5.30 P. M., we found : Of 24 perfect tests, the reaction time test agreed with the addition test 15 times, with the cubing test 11 times, with the nonsense syllables test 5 times. (This estimation of accuracy of the work in learning the nonsense syllables is of questionable accuracy because of the difficulty in properly estimating transpositions, etc.) ; mean variation agreed with addition 11 times, cubing 10 times, and nonsense syllables learned 8 times ; addition agreed with cubing 12 times, and 4 times with nonsense syllables ; cubing and nonsense agreed 11 times. On no day did all the tests agree in results, though on one day with one person all tests but the variation in reaction time agreed in showing better condition at 5.30 than at 8.30, and on another day, with one person, all tests except the nonsense syllables agreed again in showing better condition late in the afternoon.

Taking each form of test separately and comparing 8.30 A. M.

and 5.30 P. M. tests, we found : Of 24 tests, reaction time was longer in the afternoon only 10 times, variation greater only 11 times, total number of figures added less only 7 times, number added correctly less only 9 times, number of cubes written less 7 times, number of cubes written correctly less 5 times, number of nonsense syllables learned less 9 times.

This would seem to indicate a better condition at 5.30 P. M., after a day of hard study and worry, with one or two three-hour examinations often added, than at 8.30 A. M., after breakfast and a short walk. In some cases there could be no doubt that the subjects were very severely fatigued in the afternoon, though it was not often shown by the tests. One of the best records made by one of the students came just after a long, hard examination for which the student, after working and worrying all the day before, had continued work till two o'clock the night before the test.

III.

With the purpose of testing the matter with other subjects and of introducing another method, which offered more interesting matter to work with, the experiments were repeated a year later with 7 subjects, during the week before the March examinations, the week of the examinations, and the week afterwards — this last week came after three consecutive days of rest from all school work.

The tests employed were exactly the same as the year before with the addition of the combination method of Ebbinghaus, *i. e.*, filling in blanks which were made by omitting letters and small words from printed matter. The material for this test was made as follows : selections from Hamerton's *Essays on Human Intercourse* were printed with letters left out at intervals, sometimes one, sometimes enough to make a whole word. A short dash indicated the position of each missing letter. Earnest effort was made so to cut out these letters as to make it as difficult to fill the blanks in one group of lines as in another. In this we failed miserably, though half a dozen different pieces of literature were worked with before we settled on the monotonous Hamerton. It is simply impossible to make material of even approximately uniform difficulty.

If the test lasted half an hour or more, the difficulties of the lines used at each period would probably sum up about the same, but in short tests this is far from the case.

A comparison of the results of the tests taken at 8.30 A. M. and 5 P. M. during the first week, gives practically the same results as those just recorded for the previous year. The agreements and disagreements between the results of the different tests were sometimes more, sometimes less, never much over fifty per cent., often much below it. The only result which seemed even partially favorable was the learning of less nonsense syllables in the afternoon 10 out of 14 times.

Considering the results of the new test of filling in the blanks in the reading matter, we found that out of 14 perfect records more blanks were filled in the afternoon 9 times ; this test agreed in results with the reaction time test 9 times, variation in reaction time 7 times, adding test 3 times, cubing test 6 times, nonsense syllables test 7 times, when total amount of work was considered. When only the accurate work was considered in each case, the results were equally hopeless. In no instance did all the tests agree.

During the next week a change in the method of taking the reaction time was adopted, so that a choice of two reactions was made necessary instead of the simple reaction. This was done because of the utter lack of reliability in the simple reactions. Five reactions were taken at each sitting as before.

Of 24 perfect tests during this week the same lack of agreement between tests, and the same failure to indicate fatigue in the afternoon were found as before. Reaction time and variation did agree 16 times, and variation and cubing agreed 16 times, but reaction time was shorter in the afternoon than in the morning 16 times. Only 10 times out of 24 was the number of nonsense syllables learned less in the afternoon during this week, as against 10 out of 14 times the week before ; while 15 out of 24 times there were fewer blanks filled in the afternoon this week against 5 out of 14 the week before. The other results were utterly irreconcilable. Never did all tests agree.

The third week's test showed practically the same results. The reaction time was greater in the afternoon 18 out of 25

times this week, and agreed with the variation 15 out of 25 times; the number of nonsense syllables learned was greater in the afternoon 11 times out of 25. The number of blanks filled was greater in the afternoon 16 times — almost the exact opposite of the results of the previous week.

Thinking that 8.30 A. M might be too early, and the students not yet "warmed up" for the day, three subjects were now tested between 9 and 10 A. M. and 5 and 6 P. M. each day; one subject for four days, one for three, and one for two. The same tests were used, and results equally inconsistent with each other and with the undoubted facts were obtained.

It was perfectly plain that when used under exceedingly favorable conditions, by careful experimenters and with intelligent and earnest subjects, each and all of these tests were utterly unreliable as measures of fatigue in cultivated adults.

IV.

With a view of determining whether they could measure the condition of untrained children better, a new set of experiments was tried with sixth grade pupils as subjects. It was thought that possibly the habit of forced concentration and the cultivated will of a college student, would enable him to work well for the short period of the tests in spite of fatigue, while the younger pupils might not be able thus to hide their condition.

Ten children of average ability, ranging from 11 to 15 years of age, were tested for three days at 9 A. M. and between 2.45 and 3 P. M., the opening and closing hours of their school. A vacant room in the school building was used. The experiments were made by us but with the co-operation and entire sympathy of the principal of the school. The pupils entered into the tests eagerly and with remarkable earnestness. The conditions seemed ideal. The adding, cubing, learning nonsense syllables, and filling blanks were the tests used. Special pieces of easy reading were prepared for them, and the story contained in the printed matter read to them each time just before the test was made, thus making it a test of memory as well as of other mental powers.

Out of all these tests, once only, and with one boy, did all

the tests show the same thing—in this case they showed greater power at 2.45 P. M. than at 9 A. M. The failure of agreement in the tests was about the same as with adults. For instance, the addition test agreed with the cubing test 13 out of 27 times, with the filling blanks test 14 out of 29 times ; the cubing test agreed with the nonsense syllables test 15 out of 27 times, and with the filling blanks test 14 out of 26 times. The accuracy in addition and in filling blanks agreed 12 out of 29 times.

The failure to indicate fatigue in the afternoon was even more pronounced than with adults. Of 30 tests, less figures were added in the afternoon only 9 times, less columns were added correctly only 10 times. Less cubes were written only 3 out of 27 times ; the number of nonsense syllables correctly remembered was less only 10 out of 30 times ; and the number of blanks filled was less only 9 out of 29 times.

V.

A further test has this year been made with the ergograph, repeating a part of the work done by Keller.¹ Keller tested a school boy with the ergograph at frequent intervals during several days, giving him mental work between the tests. The mental work consisted in pronouncing lists of German words, and some rows of figures, prepared for this purpose. The records of two days tests are as follows :

FIRST DAY.

8	A. M.	Ergograph showed power equal to	.9776 Kgm.	
		Read 1,386 words, average time per word		.3515 sec.
8.35		Ergograph showed power equal to	1.491 Kgm.	
		Short pause. Read 1,257 words, average time per word,		.338 sec.
8.50		Ergograph showed power equal to	1.8632 Kgm.	
		Short pause. Read 425 words and some figures, average time per word		.354 sec.
9.05		Ergograph showed power equal to	1.299 Kgm.	
		Pause. Subject walking.		
10.15		Ergograph showed power equal to	.817 Kgm.	
		Read 1,364 words, average time per word		.358 sec.

¹*Biologisches Centralblatt*, XIV, Nos. 1, 2, 9.

10.35	Ergograph showed power equal to Pause. Read 399 words and some figures, average time per word	1.857 Kgm.	.349 sec.
10.50	Ergograph showed power equal to Short pause. Read 1,217 words, average time per word,	1.6962 Kgm.	.326 sec.
11.03	Ergograph showed power equal to Pause.	.864 Kgm.	
12.15 P. M.	Ergograph showed power equal to Pause.	.8298 Kgm.	
3.20	Ergograph showed power equal to Read 1,396 words, average time per word	1.0466 Kgm.	.305 sec.
3.47	Ergograph showed power equal to Read 1,266 words, average time per word	1.63 Kgm.	.283 sec.
4.05	Ergograph showed power equal to Pause.	2.156 Kgm. 1.088 Kgm.	
5.20	Ergograph showed power equal to	.8657 Kgm.	

SECOND DAY.

8.10 A. M.	Ergograph showed power equal to Read 1,392 words in 7.5 minutes, average time per word	1.4294 Kgm.	.328 sec.
8.30	Ergograph showed power equal to Short pause. Read 1,270 words in 7 minutes, average time per word	1.7206 Kgm.	.324 sec.
8.45	Ergograph showed power equal to Read 1,267 words and some figures in 8 minutes, average time per word	2.2568 Kgm.	.324 sec.
9.00	Ergograph showed power equal to Pause.	1.2436 Kgm.	
10.15	Ergograph showed power equal to Read 2,457 words in 14.5 minutes, average time per word	.7888 Kgm.	.346 sec.
10.30	Ergograph showed power equal to Pause 10 minutes. Read 2,501 words in 14-15 minutes, average time per word	1.2752 Kgm.	.356 sec.
11.00	Ergograph showed power equal to Pause.	.785 Kgm.	
12.00 M.	Ergograph showed power equal to	.3838 Kgm.	

The blind infatuation for this ergograph test is clearly shown here. It would seem much more reasonable to think that

one's ability to recognize and to call out words would come nearer indicating the condition of his mind than would one's ability to crook automatically one finger; yet Keller ignores the evidence of the time required to pronounce the words, and draws his conclusions entirely from the ergograph records.

He decides, for instance, that there must be great exhaustion during the periods from 10.50 to 11.03 A. M., and 12.15 P. M., on the first day, because the power, as shown by the ergograph, fell from 1.692 Kgm. to .864 Kgm. in 13 minutes; yet the average time required for recognizing and calling a word was not so great between 10.50 and 11.03, as between 10.35 and 10.50. If either of these methods is an accurate test of fatigue, then children are in better condition in the afternoon than in the morning, for at every test during the afternoon the words were called in less time than at any morning test, and the greatest power shown by the ergograph was at about 4 P. M.

Such unguarded experimentation really shows nothing. Crooking a finger to exhaustion every 15 minutes is a painful procedure, and no average boy will do it with uniform effort, whether fresh or fatigued. The fact that twice the ergograph results vary practically fifty per cent. in consecutive tests only 13 minutes apart, should suggest the unreliability of such a test as a measure of mental fatigue. The effects of practice, of local soreness and local fatigue, and of the changing mental attitude of the subject were not considered as they should have been.

In order to find whether the mental work had anything to do with the irregularities of this irregular ergograph record during the day, a boy 14 years old (the same age as Keller's subject), was tested by us for two days and was paid to sit around the laboratory and do nothing between times. He was naturally lazy and used to the house. The laboratory was large and comfortable, and he was free to walk around and lounge at his pleasure. He knew us well and was quite at his ease; he did as near nothing, mentally and physically, as one well could. The record is as follows :

FIRST DAY (MIDDLE FINGER, RIGHT HAND).

8.15 A. M. Ergograph test, weight lifted 78 inches.

8.50	"	"	"	"	107	"
9.05	"	"	"	"	74	"

Short walk.

10.35	Ergograph test, weight lifted 140 inches.			
10.55	"	"	"	89 "
11.10	"	"	"	80 "
11.23	"	"	"	102 "
12.30 P. M.	"	"	"	153 "

SECOND DAY (FOREFINGER).

8.35 A. M.	Ergograph test, weight lifted 64 inches.			
8.55	"	"	"	70 "
9.10	"	"	"	59 "
9.25	"	"	"	53 "

Walked several blocks.

10.40	Ergograph test, weight lifted 66 inches.			
10.55	"	"	"	48 "
11.10	"	"	"	40 "
11.25	"	"	"	46 "

Walked several blocks.

12.25 P. M.	Ergograph test, weight lifted 65 inches.			
-------------	--	--	--	--

The weight used was five pounds. The middle finger of the right hand was used the first day, and the forefinger the second day. The change was made because of the difficulty in properly adjusting the hand for the use of the middle finger.

The most noticeable things in these records are : 1st. The power, as indicated by the distance through which the weight was lifted, varied all the way from 74 inches to 153 inches on the first day, and from 46 inches to 70 inches on the second day; 2nd. The tests after the hour intermission and walk, and the last test before dinner on each day, showed very great power. The variations were wide and irregular, without any work being done on either day.

On another day four boys were tested by the method employed by Smedley, in Chicago.¹ Two were required to work arithmetic examples between the tests, while two did nothing but walk around the building and lounge in the laboratory. Both of the boys who did not work showed a gain of power for the first three tests, but both showed a falling off at 12.20, the last test of the morning, and both showed slightly less power then than at 9 o'clock. One of the boys, who worked all the

¹Report on Child Study Investigation, Annual Report Board of Education of Chicago, 1898-'99.

time, improved regularly until 11.30, and showed only a slight falling off at 12.20, although his record at that time was higher than at either of the first two tests of the morning. The record of the other boy who worked was marred by poor adjustment of the hand in the last test ; but at the 11.30 test he was just as strong as at the first test. All of them showed a falling off in the afternoon except one working boy who showed a slight gain of power the first test after dinner, but a marked decrease in the later tests. The other three boys did not show very much loss of power in any of the afternoon tests. They were all quite tired of (or bored with) the experiment by this time and were anxious to get out. They really did not seem to try so hard after dinner as in the morning.

VI.

After three years of periodic trial of all these methods of testing fatigue, the results force upon us the conclusion that they are, as used at present, worthless as tests of fatigue either in children or in adults. Two things further seem to us quite clear : 1st, That no such test extending over so brief a period as two to five minutes can be accurate in measuring fatigue, because any average student can, by a strong effort of the will, practically nullify for such a period the effect of any fatigue that is not pathological ; 2nd, On the other hand, the tests continuing for an hour or more, using uniform work, such as addition, are of necessity made with material that is not uniform, or is uniformly stupid and uninteresting. No interesting work seems to be measurably uniform in difficulty or uniformly interesting.

The problems of fatigue will never be settled till such tests are abandoned and the results gotten by them recognized as worthless. When we surrender our present vain knowledge, and come to the problem fresh and unbiased, there is hope of solving it.

The first thing needed is a clearer analysis of the conditions present when one is fatigued ; now, very varied states are roughly classed as states of fatigue. When the components of fatigued states are better known, then a method must be devised, either to eliminate in turn each component or to measure

its effect. Thorndike is the only other worker in the field who has properly recognized the complexity of the fatigued states and attempted to analyze them. He, unfortunately, did not follow up this line and try to eliminate or to measure and evaluate these several components. A part of this work we have roughly planned but have yet to complete.

Before any specific state of so-called fatigue can be intelligently considered, several facts must be determined.

1st. What part of the apparent fatigue (or inefficiency), is due to purely local physiological conditions?

2nd. What part to purely mental conditions?

(a) What part is due to suggestion, conscious or unconscious?

(b) What part is merely lack of interest, or "bore?"

(c) What part is due to a pure mental fatigue?

3rd. How far can the will of the person tested go in hiding or overcoming the effects of each of the elements of fatigue, and how long can this effort be maintained?

4th. Does the test employed extend beyond this period of volitionally modified effort; and, if so, does it not itself produce the fatigue it may measure?

5th. What is the emotional attitude of the subject toward the test employed, and how far will this influence the results in each case even with conscientious subjects? Can any test be found of uniform interest to any large number of pupils?

When it is recognized that "fatigue" is itself not the simple matter it has been considered, and that the attitude of the subject towards the test largely determines the result, it will be readily recognized that the search for a single, simple, accurate test of fatigue that can be employed at any time with any number of pupils by a half-trained teacher is a vain quest.